Estimating mapped-plot forest attributes with ratios of means

S.J. Zarnoch and W.A. Bechtold

Abstract: The mapped-plot design utilized by the U.S. Department of Agriculture (USDA) Forest Inventory and Analysis and the National Forest Health Monitoring Programs is described. Data from 2458 forested mapped plots systematically spread across 25 states reveal that 35% straddle multiple conditions. The ratio-of-means estimator is developed as a method to obtain estimates of forest attributes from mapped plots, along with measures of variability useful for constructing confidence intervals. Basic inventory statistics from North and South Carolina were examined to see if these data satisfied the conditions necessary to qualify the ratio of means as the best linear unbiased estimator. It is shown that the ratio-of-means estimator is equivalent to the Horwitz–Thompson, the mean-of-ratios, and the weighted-mean-of-ratios estimators under certain situations.

Résumé : Les auteurs décrivent le dispositif de parcelles cartographiées utilisé par les programmes d'analyse et d'inventaire forestier et de suivi national de l'état de santé des forêts du Département de l'agriculture des États-Unis. Les données provenant de 2458 parcelles cartographiées en forêt et réparties dans 25 États montrent que 35% des parcelles chevauchent des conditions multiples. L'estimateur du ratio des moyennes est la méthode qui a été développée pour obtenir un estimé des caractéristiques de la forêt à partir des parcelles cartographiées de même que des mesures de variabilité utiles pour construire des intervalles de confiance. Les statistiques de base de l'inventaire de la Caroline du Nord et du Sud ont été examinées pour vérifier si ces données rencontrent les conditions nécessaires pour faire du ratio des moyennes le meilleur estimateur linéaire non biaisé. Les résultats démontrent que l'estimateur du ratio des moyennes est équivalent à l'estimateur de Horwitz–Thompson, à celui de la moyenne des ratios et à l'estimateur pondéré de la moyenne des ratios dans certaines conditions.

[Traduit par la Rédaction]

Introduction

The increased demand for forest products and the recent interest in the potential effects of atmospheric deposition and global warming on forest resources have emphasized the need for accurate forest inventories and forest health monitoring. To address these issues, the U.S. Department of Agriculture (USDA) Forest Inventory and Analysis (FIA) and the National Forest Health Monitoring (FHM) Programs use mapped-plot designs for extensive forest inventories (Hahn et al. 1995). Since these data are used by other government organizations, industry, and academia, it is imperative that the sampling methodology be clearly defined with appropriate estimators and measures of precision.

Scott and Bechtold (1995) recommend using the ratio of means (ROM) estimator to calculate inventory attributes from mapped plots. However, complications occur because each plot is subsampled with two different-sized plots for different-sized vegetation. Other complexities are caused by varying plot sizes resulting from poststratification of the data.

The objectives of this paper are to further develop the ROM estimator for processing mapped-plot data, utilize FHM inventory data to determine if the assumptions required for using the ROM estimator are satisfied for some basic inventory statistics, and discuss the relationship of the ROM estimator to alternative estimators.

Received June 21, 1999. Accepted December 9, 1999.

S.J. Zarnoch[1] and W.A. Bechtold. USDA Forest Service, Southern Research Station, P.O. Box 2680, 200 Weaver Boulevard, Asheville, NC 28802, U.S.A.

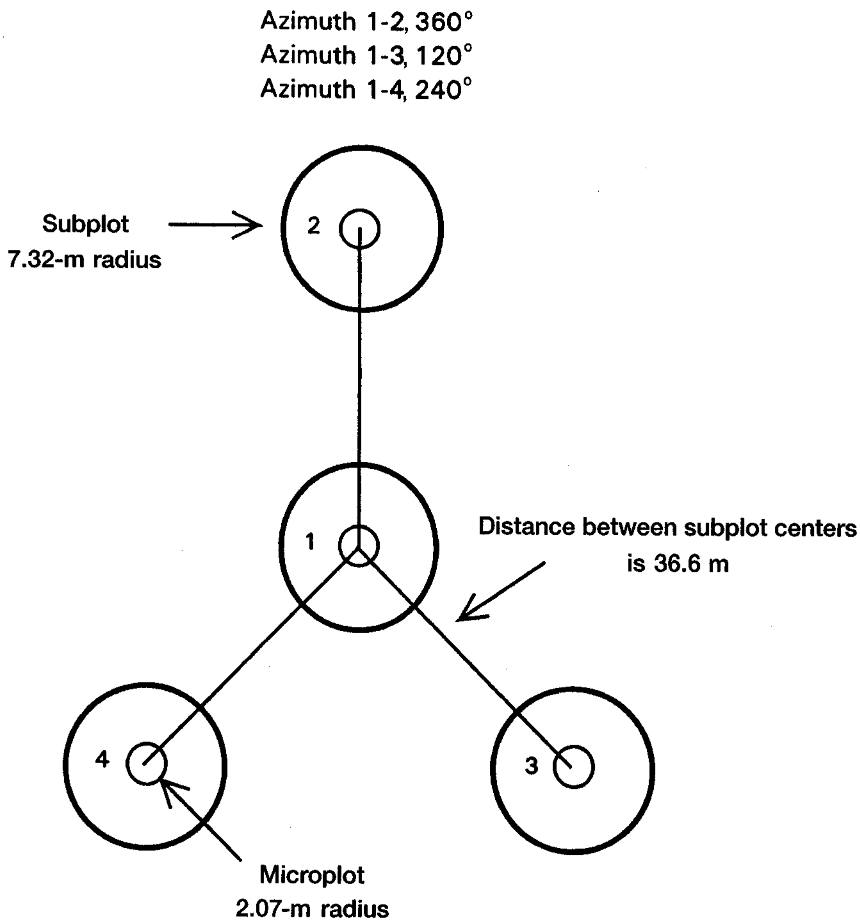
[1]Corresponding author. e-mail: szarnoch/srs_fia@fs.fed.us

Mapped-plot design

FIA and FHM inventories are comprised of a series of fixed-size ground plots from which measurements of forest attributes are obtained. FIA uses a design that is a combination of random and systematically located plots. FHM utilizes a strictly systematic sample where each plot is located on a hexagonal grid. Both programs use a ground plot that consists of a cluster of four 0.0166-ha fixed-area circular subplots (7.32-m radius) spaced 36.6 m apart (Fig. 1). Each subplot includes a 0.0013-ha fixed-area circular microplot (2.07-m radius). Trees at least 12.7-cm DBH are measured on the subplots, while saplings between 2.54- and 12.6-cm DBH are sampled on the microplots. Across each cluster, the four subplots encompass a total of 0.0665 ha, and the microplots cover 0.0053 ha.

Cluster designs are preferred for extensive inventories because they are more cost effective than single-point plots of similar size since they reduce the between-plot variation (Scott et al. 1983; Scott 1993). For all conventional inventory applications, data are summed across the four subplots and across the four microplots, and the cluster is processed as a single plot consisting of one combined subplot and one combined microplot.

Fig. 1. Mapped-plot configuration.



Mapped plots are based on the concept of "condition classes," which are assigned by field crews to classify land area into homogeneous groups. Plots are characterized by several predetermined discrete variables such as land use (forest versus nonforest), forest type, and stand origin (planted versus natural). Plots that are not homogeneous are subdivided into multiple condition classes. Survey crews assign an arbitrary number (usually 1) to the first condition class encountered on a plot. This number is defined by the series of condition-class variables attached to it. Additional condition classes are identified if there is a distinct change in any of the condition-class variables on the plot. When multiple condition classes occur on a particular subplot or microplot, field crews map the boundaries between conditions with azimuth readings (Scott and Bechtold 1995). Subplot and microplot areas in each condition class are subsequently computed for each plot when the data are processed. All trees tallied are assigned to the condition classes in which they are located. Since nonforest areas are not part of the FIA and FHM populations of interest, no tree data are recorded for nonforest land uses.

At first glance, an unwieldy number of condition-class permutations seem likely at the regional scale, especially since condition classes from the same data set must be processed in different combinations from one inventory summary table to the next (depending on the attribute of interest). However, most plots have only 1 or 2 condition classes, and data summarizations are easily managed with indicator functions, as will be demonstrated.

Ratio-of-means estimators

This section presents the ratio-of-means estimator and several variants. Much of this originates from Cochran (1977) and serves as reference material for the next section, which specifically applies ROM methodology to forest inventories using the mapped-plot design.

The attributes of interest in a forest inventory cover a broad spectrum. They might include estimates of average numbers of trees per hectare, total inventory volume in a region, mean quadratic DBH, mean number of conks per tree, proportions of forest area by forest type, etc. All of these estimates involve ratios (y/x) where y is some attribute of interest and x is some correlated auxiliary variable (usually an area or tree total). Several alternative estimators are possible, but the ROM estimator is the best linear unbiased estimator (BLUE) under the following conditions (Cochran 1977, sect. 6.7):

(1) the relationship between y_i and x_i is linear through the origin and

(2) the variance of y_i is proportional to x_i.

If only condition 1 is met, then the ROM estimator is still unbiased (Ek 1971) but is no longer BLUE (Brewer 1963; Royall 1970).

In addition, as is the case with all other estimators, the sample size should be sufficiently large to ensure a reliable estimate. Hence, Cochran (1977, sects. 6.3 and 6.8) recommends that the number of observations be at least 30 and large enough so that the coefficients of variation of $\bar{y}$ and $\bar{x}$ are both less than 10%.

Cochran (1977, sect. 6.2) defines the ROM estimator as

$$[1] \qquad \hat{R} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} = \frac{\bar{y}}{\bar{x}}$$

where y_i is the variable of interest on plot i, x_i is an auxiliary variable on plot i that is correlated with y_i, and n represents the number of plots selected from the population by simple random sampling. Although this estimator is known to be biased and positively skewed, as the sample size increases above 30 the bias becomes negligible and the distribution approaches normality. In a simulation study, Ek (1971) found that the bias of several ratio of means estimators of forestry attributes was less than 3.9% when the sample size was only four.

Cochran (1977, sect. 6.4, eq. 6.13) specifies the variance of the ROM estimator as

$$[2] \qquad \hat{V}(\hat{R}) = \frac{1}{n\left(\sum_{i=1}^{n} x_i/n\right)^2}(S_y^2 + \hat{R}^2 S_x^2 - 2\hat{R}S_{yx})$$

or, in an alternate form,

$$[3] \qquad \hat{V}(\hat{R}) = \frac{1}{n(n-1)\left(\sum_{i=1}^{n} x_i/n\right)^2} \times \left(\sum_{i=1}^{n} y_i^2 + \hat{R}^2 \sum_{i=1}^{n} x_i^2 - 2\hat{R}^2 \sum_{i=1}^{n} y_i x_i\right)$$

where S_y^2 and S_x^2 are the typical sample variances of y and x, respectively, and S_{yx} is their covariance. Note that the finite population correction has been ignored (as it will be throughout this paper) since the sampling fraction is unknown and negligible in most forest inventories. If the sample size is at least 30 and the coefficients of variation of $\bar{y}$ and $\bar{x}$ are both less than 10%, then confidence intervals may be constructed in the usual way as

$$[4] \qquad \hat{R} \pm z(\hat{V}(\hat{R}))^{0.5}$$

where z is the normal deviate at the desired confidence level. If these conditions are not met, alternative but more complex methods may be used (Fieller 1932; Paulson 1942; James et al. 1974).

In some situations (to be addressed in more detail later) it is desirable to formulate an estimator that is the sum of two correlated ROM estimators. This occurs when the estimator is derived from trees on both microplots and subplots. In this case, the estimator is defined as

$$[5] \qquad \hat{R}^* = \hat{R} + \hat{R}'$$

where $\hat{R}$ is the ratio-of-means estimator for the subplot and $\hat{R}'$ is the ratio-of-means estimator for the microplot. An alternative to eq. 5 could be formulated by first combining the subplot and microplot data on each plot and then calculating a ROM estimator from the combined data, but we rejected this approach for a couple of reasons. First, depending on the condition classes of interest, the ratio of subplot area to microplot area may not be constant from plot to plot. Simply expanding microplot data to the subplot level creates an inverse relationship between a microplot's size and the weight it receives when expanded. In effect, all microplots would lose their size identity and contribute equally to the variance. Secondly, there is also the possibility that a condition located on a subplot is not sampled by a microplot, causing $\hat{R}$ and $\hat{R}'$ to be based on a different number of plots.

The estimated variance of eq. 5, which includes the covariance between the two estimators, is

$$[6] \qquad \hat{V}(\hat{R}^*) = \hat{V}(\hat{R}) + \hat{V}(\hat{R}') + 2\hat{C}(\hat{R}, \hat{R}')$$

where Cochran (1977, sect. 6.18, eq. 6.90) defines the covariance term as

$$[7] \qquad \hat{C}(\hat{R}, \hat{R}') = \frac{1}{n(n-1)\left(\sum_{i=1}^{n} x_i/n\right)\left(\sum_{i=1}^{n} x_i'/n\right)} \times \left(\sum_{i=1}^{n} y_i y_i' - \hat{R}\sum_{i=1}^{n} y_i' x_i - \hat{R}'\sum_{i=1}^{n} y_i x_i' + \hat{R}\hat{R}'\sum_{i=1}^{n} x_i x_i'\right)$$

For other situations (to be addressed later), useful estimators can be formed as ratios of two correlated ROM estimators. This estimator is defined as

$$[8] \qquad \hat{R}^{**} = \frac{\hat{R}_1^*}{\hat{R}_2^*}$$

with estimated variance (Cochran 1977, Sec. 6.19, eq. 6.95)

$$[9] \qquad \hat{V}(\hat{R}^{**}) = \hat{R}^{**2}\left(\frac{\hat{V}(\hat{R}_1^*)}{\hat{R}_1^{*2}} + \frac{\hat{V}(\hat{R}_2^*)}{\hat{R}_2^{*2}} - \frac{2\hat{C}(\hat{R}_1^*, \hat{R}_2^*)}{\hat{R}_1^*\hat{R}_2^*}\right)$$

where all quantities are as previously defined and

$$[10] \qquad \hat{C}(\hat{R}_1^*, \hat{R}_2^*) = \hat{C}(\hat{R}_1 + \hat{R}_1', \hat{R}_2 + \hat{R}_2') = \hat{C}(\hat{R}_1, \hat{R}_2) + \hat{C}(\hat{R}_1, \hat{R}_2') + \hat{C}(\hat{R}_1', \hat{R}_2) + \hat{C}(\hat{R}_1', \hat{R}_2')$$

Estimating means of forest attributes

So that classical sampling methodology can more easily be applied, it is assumed the sample of plots is a simple random sample where the sampling unit is the plot, which consists of one combined subplot and one combined microplot. When the sample is partially systematic (FIA) or totally systematic (FHM) this assumption may be challenged owing to the spatial pattern of trees across the forest landscape. However, it is a reasonable assumption for most forest inventories that systematic sampling is approximately equivalent to simple random sampling. Milne (1959), who studied spatial periodicity in forest populations, states that “the danger to centric systematic sampling from unsuspecting periodicity is so small as to be scarcely worth a thought."

Separate subplot or microplot estimates

In some cases the attribute of interest can be obtained either from the microplot data or from the subplot data. In other situations it is necessary to combine subplot and microplot data. An example of the former would be analyses that require only trees 12.7-cm DBH and larger; an example of the latter would be analyses involving trees 2.54-cm DBH and larger. The application of the ROM estimator to combined estimates is slightly more complex, so the separate case is addressed first.

For convenience, all attributes commonly estimated from inventory data can be grouped into three general cases:
(1) per-hectare-level attributes[2] (average trees per hectare, basal area per hectare, etc.),
(2) tree-level attributes (average DBH, number of conks per tree, etc.), and
(3) stand-level attributes (average stand age, percent stocking, etc.).

The ROM estimator is appropriate for attributes in all three categories provided y_{ij} and x_{ij} as described below are properly defined. For a given plot i let

y_{ij} = the value of the y variable of interest on plot i and condition class j

x_{ij} = the value of the x variable of interest on plot i and condition class j

c_i = the number of condition classes on plot i

I_{ij} = 1 if condition class j on plot i is of interest or

= 0 if condition class j on plot i is not of interest

n = number of plots containing at least one condition class of interest.

Then, the indicator function I_{ij} is used to obtain the values for y_i and x_i on plot i as

$$y_i = \sum_{j=1}^{c_i} I_{ij} y_{ij}$$

and

$$x_i = \sum_{j=1}^{c_i} I_{ij} x_{ij}$$

Examples of how y_{ij} and x_{ij} are defined for each of the three general cases follow:

Case 1. Per-hectare-level attribute (e.g., average basal area per hectare)

y_{ij} = sum of the basal areas of all trees tallied on plot i in condition class j

x_{ij} = area of plot i in condition class j

Case 2. Tree-level attribute (e.g., average DBH)

y_{ij} = sum of the DBH's for all trees tallied on plot i in condition class j

x_{ij} = the number of trees tallied on plot i in condition class j

Case 3. Stand-level attribute (e.g., average stand age)

y_{ij} = (subplot area of plot i in condition class j) × (stand age of plot i in condition class j)

x_{ij} = subplot area of plot i in condition class j

Note that case 3 is similar to a traditional weighted mean where the weight is the area of plot i in condition class j (i.e., the x_{ij}). In fact, the mean estimates will be identical, but the variances will differ, since the traditional weighted mean treats the weights as constants when calculating the variance, while the ratio of means correctly considers them as variables.

It is also interesting to observe that per-hectare-level estimates (case 1) can be formulated with the stand-level approach (case 3) if y_{ij} is redefined. For instance, the basal area example from case 1 above conforms to case 3 if y_{ij} is redefined as

y_{ij} = (area of plot i in condition class j) × (basal area per hectare of plot i in condition class j)

= (x_{ij}) × (basal area on plot i and condition class j/x_{ij})

= basal area on plot i and condition class j.

Thus, the y_{ij}'s are identical for both approaches. The per-hectare-level and stand-level cases are equivalent if y_{ij} is properly formulated, hence the user may select the most intuitive method.

Combined estimates

When estimates are needed for attributes that require data from both the microplot and subplot (e.g., when an estimate is based on trees 2.54-cm DBH or larger), three different approaches are recommended depending on whether the attribute is case 1, case 2, or case 3.

To combine per-hectare-level (case 1) estimates from the subplots and microplots, the individual estimates are simply added as specified in eq. 5, where $\hat{R}$ and $\hat{R}'$ are the ratio-of-means estimates for the subplot and microplot components, respectively. The variance can then be obtained from eq. 6, where x_i and y_i are data from the subplots, and x_i' and y_i' are data from the microplots.

In contrast with per-hectare attributes, estimators for tree-level attributes (case 2) cannot be obtained by combining subplot and microplot data. For instance, the average DBH of all stems is not mean sapling DBH plus mean tree DBH. The correct result can be obtained by creating two additive combined-ratio estimates, one that expresses the tree-level

[2] Note that the per-hectare-level values yield population totals (total number of trees or basal area in a region) when multiplied by the appropriate population area. This is addressed in more detail later.

attribute on a per-hectare basis $\hat{R}_1^*$ and the other to estimate trees per hectare $\hat{R}_2^*$. Then a ratio of these estimates is formulated as in eq. 8 with estimated variance obtained from eq. 9.

Stand-level (case 3) attributes produced from combined microplot and subplot data pose more difficulty. Such variables usually involve some type of index that requires the microplot and subplot data to be combined before the index can be produced, so it would be inappropriate to add or average subplot and microplot indices that are computed separately. Examples might include species diversity index, stand density index, or stocking.

Scott and Bechtold (1995) recommend expanding the microplot data used in the calculation of such indices by a weight defined as the subplot/microplot area ratio. In effect, this puts the microplot data on the same basis as the subplot for each condition of interest, and yields a plot value defined as

$$z_i = \left(\frac{x_i}{x_i'}\right)y_i' + y_i$$

The mean of z_i and its variance can then be obtained in the same manner as a separate stand-level estimator (e.g., stand age), but this estimator has some noteworthy disadvantages. First, the size of each subplot and microplot depends on the condition classes of interest as determined by the indicator function I_{ij}, so the weights can vary from plot to plot. This inflates the variance, because the smallest microplots (which are usually the most variable) receive the most weight since their subplot/microplot ratios are the largest. Secondly, it does not function well for plots where the condition class of interest is present on the subplot, but not the microplot. In this case subplots containing valuable data will be discarded whenever corresponding microplots have no data.

Estimating area

Proper formulation of proportions involving area estimates is crucial to processing mapped-plot data. All area estimates must conform to a specific base line area, which is defined by the estimation problem at hand. For example, the proportion of loblolly pine forest types in the geographic area of interest might be the ratio of loblolly pine area to "total" land area (forest plus nonforest), or it might be the ratio of loblolly pine to "forest" land area in the region of interest.

The composition of the base line area thus defines the condition class that becomes the denominator of all ratios involving area. An estimator for such a proportion is

$$[11] \quad \hat{P}_C = \frac{\sum_{i=1}^{n_B} x_i}{\sum_{i=1}^{n_B} x_{Bi}}$$

where $\hat{P}_C$ = ratio-of-means estimator for the proportion of area in the condition classes of interest with respect to the base line, x_i = area of plot i that is in the condition classes of interest, x_{Bi} = area of plot i which is in the base line condition class, and n_B = the number of plots in the base line condition class.

Note that the x_i and x_{Bi} must be obtained from the plot data by use of two different indicator functions, each defined by their respective condition classes. The sample variance of $\hat{P}_C$ would be obtained from an equation analogous to eqs. 2 or 3.

If the total area in the base line is known, then an estimate of total area in the condition class of interest is

$$[12] \quad \hat{A}_C = A_B\hat{P}_C$$

where $\hat{A}_C$ is the total area in the condition classes of interest and A_B is the total area (known) in the base line. An estimate of the variance is

$$[13] \quad \hat{V}(\hat{A}_C) = A_B^2\,\hat{V}(\hat{P}_C)$$

If A_B is not known, but estimated from a sample that is independent of that used to obtain $\hat{P}_C$, then

$$[14] \quad \hat{A}_C = \hat{A}_B\hat{P}_C$$

and based on Kish (1965)

$$[15] \quad \hat{V}(\hat{A}_C) = \hat{A}_B^2\,\hat{V}(\hat{P}_C) + \hat{P}_C^2\,\hat{V}(\hat{A}_B)$$

where $\hat{V}(\hat{A}_B)$ is obtained by whatever method is appropriate for the independent sample.

Expanding mean per-hectare estimates to population totals

One of the most basic objectives of a forest survey is to estimate population totals (e.g., total basal area or volume in a condition class of interest for a given geographic unit). To obtain population totals, the ROM per-hectare-level estimator ($\hat{R}$) must be multiplied by the total area in the condition class of interest (A_C) in the geographic unit. There are several ways to obtain this estimate, depending on the level of information available. A_C may be known, it may come from an independent estimate such as aerial photography, or it may be estimated directly from the mapped-plot data.

Kish (1965, page 211) recommends the following general formula to calculate the variance of estimators resulting from the product of two random variables (X and Y):

$$[16] \quad V(XY) = X^2V(Y) + Y_2V(X) + 2XYC(X,Y)$$

Depending on the source of A_C, some of the terms in eq. 16 may drop out as discussed below.

A_C is known

In the ideal situation, A_C is known without error, resulting in an estimator with maximum precision. This type of information may be available for private landowners with small, intensely managed forests. When A_C is known, the separate population total estimator is defined as

$$[17] \quad \hat{T} = A_C\hat{R}$$

Since A_C is known, $V(A_C) = 0$, and A_C and $\hat{R}$ are not correlated ($C(A_C,\hat{R}) = 0$), so eq. 16 yields the variance

$$[18] \quad \hat{V}(\hat{T}) = A_C^2\,\hat{V}(\hat{R})$$

When total estimates for a combined microplot and subplot attribute are needed, $\hat{R}^*$ is used instead of $\hat{R}$.

A_C is estimated

Usually A_C is not known, because the land area has not been surveyed or there have been changes to A_C over time owing to ecological succession, natural disturbances, harvesting practices, etc. If $\hat{A}_C$ is estimated from a source independent of the mapped plots (e.g., photography or satellite imagery), then the separate population total estimator is

$$[19] \quad \hat{T} = \hat{A}_C\hat{R}$$

and the sample variance is

$$[20] \quad \hat{V}(\hat{T}) = \hat{A}_C^2\,\hat{V}(\hat{R}) + \hat{R}^2\,\hat{V}(\hat{A}_C)$$

Since there is an increase in uncertainty due to a lower level of information, there is an increase in the variance of the estimator. Again, for combined estimates, $\hat{R}^*$ is used instead of $\hat{R}$.

A_B is known

If A_C is not known, and no independent estimate is available, but some base line A_B is known (e.g., total land area in the region of interest), then the subplot[3] data could be used to estimate $\hat{A}_C$ (eq. 12) along with $\hat{R}$. The separate population total estimator is then defined as

$$[21] \quad \hat{T} = A_B\hat{P}_C\hat{R} = \hat{A}_C\hat{R}$$

The sample variance must incorporate the covariance between $\hat{A}_C$ and $\hat{R}$, and is thus derived from eq. 16 as

$$[22] \quad \hat{V}(\hat{T}) = \hat{A}_C^2\hat{V}(\hat{R}) + \hat{R}^2\hat{V}(\hat{A}_C) + 2\hat{A}_C\hat{R}\hat{C}(\hat{A}_C, \hat{R})$$

This estimator may be more or less precise than the previous one, depending on the precision of the $\hat{A}_C$ estimates and the covariance term. Note that the first population estimator (eq. 17) is a special case of this estimator if A_B is equal to A_C (which implies that $\hat{P}_C = 1$).

Although the variance eq. 22 could be computed from previous techniques presented in this paper, the mathematics are cumbersome. A simplification is possible if it is noted that the estimator (eq. 21) can be written as

$$[23] \quad \hat{T} = A_B\frac{\sum_{i=1}^{n_B} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n_B} x_{Bi} \sum_{i=1}^{n} x_i}$$

where there are n_B plots in the base line and n plots in the condition class of interest which define the upper bounds for the summations in eq. 23. Note that

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n_B} y_i$$

and

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n_B} x_i$$

since y_i and x_i are defined with respect to the condition class of interest and, hence, the $n_B - n$ plots not in the condition class of interest have $y_i = 0$ and $x_i = 0$. Thus, substituting these relationships into eq. 23 and canceling terms, the estimator becomes

$$[24] \quad \hat{T} = A_B\frac{\sum_{i=1}^{n_B} y_i}{\sum_{i=1}^{n_B} x_{Bi}} = A_B\hat{R}_B$$

where $\hat{R}_B$ is defined as the ratio-of-means estimator with respect to the base line using the x_{Bi} instead of the typical x_i. The variance is then obtained as

$$[25] \quad \hat{V}(\hat{T}) = A_B^2\hat{V}(\hat{R}_B)$$

where $\hat{V}(\hat{R}_B)$ can be calculated in a fashion similar to eqs. 2 or 3. For combined estimates, individual estimates are computed and summed, and the variance is computed in a fashion analogous to eq. 6.

A_B is estimated

Instead of estimating the area in each specific condition class of interest $\hat{A}_C$ as required for eq. 19, it may be advantageous to derive an estimator that utilizes an estimate of the area in some base line ($\hat{A}_B$), such as the total area of forest in the region of interest. This case would apply to FIA when two-phase sampling is employed where total area of forest is estimated from photography and ground plots are used to partition the forest area. If $\hat{A}_B$ is estimated from a source that is independent of the mapped plots, then the separate population total estimator is

$$[26] \quad \hat{T} = \hat{A}_B\hat{R}_B$$

where $\hat{R}_B$ is again defined as the ratio of means estimator with respect to the base line, and the sample variance is

$$[27] \quad \hat{V}(\hat{T}) = \hat{A}_B^2\hat{V}(\hat{R}_B) + \hat{R}_B^2\hat{V}(\hat{A}_B)$$

Tree expansion factors

For data storage and retrieval purposes it is sometimes convenient to derive a tree weight (or expansion factor), which is the amount that each individual sample tree contributes to the estimate of the population total. With the ROM estimator, the tree weights are constant with respect to whatever base line is used to produce the total. If A_C (or estimated $\hat{A}_C$) is used, the tree weight is dependent on the condition class A_C (or $\hat{A}_C$). From eq. 17, the expansion factor (E) assigned to all trees in condition A_C is

$$E = \frac{A_C}{\sum_{i=1}^{n} x_i}$$

Estimated $\hat{A}_C$ is substituted for A_C in the above when computing the expansion factor associated with eq. 19.

[3] Although it is possible to estimate area with the microplot data, the subplot data are recommended for this purpose since more area is sampled.

If A_B (or $\hat{A}_B$) is used, the tree weight is constant with respect to the base line A_B, but independent of condition class. From eq. 21 it appears that

$$E = \frac{\hat{A}_C}{\sum_{i=1}^{n} x_i}$$

which is condition class dependent. However, since $\hat{A}_C = A_B\hat{P}_C$ and

$$\hat{P}_C = \frac{\sum_{i=1}^{n_B} x_i}{\sum_{i=1}^{n_B} x_{Bi}}$$

we have, upon substituting

$$E = \frac{\hat{A}_C}{\sum_{i=1}^{n} x_i} = \frac{A_B\hat{P}_C}{\sum_{i=1}^{n} x_i} = \frac{A_B}{\sum_{i=1}^{n} x_i} \frac{\sum_{i=1}^{n_B} x_i}{\sum_{i=1}^{n_B} x_{Bi}} = \frac{A_B}{\sum_{i=1}^{n} x_i} \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n_B} x_{Bi}}$$
$$= \frac{A_B}{\sum_{i=1}^{n_B} x_{Bi}}$$

which is condition class independent. Thus, when the ground plots are used to estimate the area in a particular condition from some base line, the tree expansion factor is a constant for all condition classes within the specified base line area A_B (or $\hat{A}_B$).

Data analysis

FHM plot data were obtained for 25 states. All plots were systematically located and mapped according to six condition-class variables: land use, forest type, stand size, stand origin, past disturbance, and density class. Since variability of plot size is one of the properties upon which the ROM estimator is dependent, distributions of plots with multiple conditions were evaluated to determine if there is enough forest fragmentation to justify an estimator that accommodates variable plot sizes. A subset of these plots from North Carolina and South Carolina were further checked to determine if the mapped inventory data met the prerequisites for making the ROM a candidate for best linear unbiased estimator.

Mapped-plot condition-class distribution

Data from 25 states show that 35% of all forested plots (the population of interest for FHM and FIA) are less than full size because they contain two or more condition classes (Table 1). High percentages of partial plots underscore the need for an estimator that accounts for varying plot size. High percentages of plots with multiple conditions also support the general use of the mapped design, which properly stratifies the excessive fragmentation observed in these data. Under previous designs, fragmented plots would have been "fuzzed" or "rotated" into a homogenous condition (Hahn et al. 1995).

Best linear unbiased estimator

Trees per hectare is arguably the most basic inventory statistic produced by FIA. All other tree attributes (e.g., volume per hectare, basal area per hectare, conks per hectare, etc.) are likely to exhibit the same statistical properties as number of trees per hectare. Data from North and South Carolina were used to check this inventory attribute to see if it met the conditions necessary for the ROM to be the best linear unbiased estimator. The bivariates y_i (subplot trees tallied) and x_i (subplot area sampled) were obtained for each forested condition class sampled in these two states. Several descriptive statistics for y_i and x_i are shown in Table 2. Plot size (y_i) ranged from small pieces of a total plot 0.0166 ha (1/4 of a plot) to 0.0673 ha (a full plot). Scatterplots of the y_i and x_i values show a linear trend through the origin with the variance approximately proportional to x_i (Figs. 2 and 3). The linear relationship between the bivariates is strong, with correlation coefficients of 0.38 and 0.31 for North Carolina and South Carolina, respectively (Table 2). The assumptions were further tested by fitting linear regressions to the x_i and y_i and estimating the intercept and slope coefficients. For both states, tests of the hypothesis that the intercept is equal to zero could not be rejected, and the slope coefficients are highly significant. In addition, the coefficients of variation of the means are less than 10% in all cases (Table 2).

Discussion

Alternative estimators

Other estimators could be used with mapped-plot inventory data, but these are usually inferior to the ROM when the BLUE conditions are met. Compared with alternatives, ROM also has several practical advantages, as discussed below.

Horwitz–Thompson (HT)

The HT estimator (Cochran 1977, Sec. 9A7, eq. 9A.37) is often used in forest inventory to estimate population totals. Applied to the mapped design, the HT estimator is defined as

$$\hat{T}_{\mathrm{HT}} = \sum_{i=1}^{n} \sum_{j=1}^{c_i} \sum_{k=1}^{m_{ij}} \frac{I_{ij} y_{ijk}}{\pi_{ijk}}$$

where π_{ijk} = probability that tree k in condition class j on plot i is selected by the sampling method, and m_{ij} = number of trees in condition class j on plot i.

An estimator for the true π_{ijk} could be defined as the proportion of the total area sampled in condition class j. Thus,

$$\hat{\pi}_{ijk} = \hat{\pi} = \sum_{i=1}^{n} \sum_{j=1}^{c_i} \frac{I_{ij} x_{ij}}{A}$$

where A is the area over which the estimated total is desired. Given this, it can then be demonstrated that the HT estimator

Table 1. Number of forested FHM plots, and the percentage with multiple condition classes, by State.

State	No. of forested plots	% with multiple conditions
Alabama	131	46
California	190	26
Colorado	150	39
Connecticut	11	27
Georgia	162	43
Idaho	142	24
Illinois	40	58
Indiana	37	38
Maine	127	18
Maryland	45	44
Massachusettes	22	32
Michigan	130	42
Minnesota	256	36
New Hampshire	34	47
New Jersey	13	54
North Carolina	111	45
Oregon	195	20
Pennsylvania	80	34
South Carolina	81	35
Vermont	24	25
Virginia	101	44
Washington	136	27
West Virginia	77	40
Wisconsin	95	46
Wyoming	68	26
Total	2458	35

Note: Forested plots have at least one condition class in a forest land use.

is equivalent to the ROM estimator, since substituting this value for $\hat{\pi}_{ijk}$ yields

$$\hat{T}_{\mathrm{HT}} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{c_i}\sum_{k=1}^{m_{ij}} I_{ij}y_{ijk}}{\sum_{i=1}^{n}\sum_{j=1}^{c_i} I_{ij}x_{ij}/A} = A\frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} = A\hat{R}_{\mathrm{ROM}}$$

Although these estimators are equivalent, several difficulties are encountered when estimating the variance associated with the HT estimator as applied to mapped plots. First, the HT estimator defines individual trees as the sampling units for the HT estimator, but the clustered nature of the trees is not accounted for in the variance. Secondly, the selection probabilities are estimated from the data and are not known, fixed quantities. The application of a common π to all the trees is questionable, because the random nature of π does not enter into the variance. Thirdly, the plot sizes themselves are random variables, but HT ignores this source of variation. The ROM estimator circumvents these problems by utilizing the plot as the basic sampling unit.

Mean of ratios (MOR)

The mean of ratios (MOR) estimator is often recommended as an alternative to the ratio of means when the

Table 2. Descriptive statistics and the linear relationship for the x and y variables used to estimate trees per hectare from the FHM plots in North Carolina and South Carolina.

(A) Descripive statistics for x and y variables

Statistic	North Carolina x	North Carolina y	South Carolina x	South Carolina y
No. of plots	111	111	81	81
Mean	0.0601	22.8	0.0621	21.9
Minimum	0.0168	0	0.0168	0
Maximun	0.0673	73	0.0673	66
Standard deviation	0.0135	14.8	0.0118	16.7
Coefficient of variation	22.5	65.0	19.0	76.2
Coefficient of variation of the mean	2.14	6.17	2.11	8.46

(B) Linear relationship between x and y

	North Carolina	South Carolina
Correlation coefficient	0.38	0.31
Intercept estimate B_0	−1.88	−5.07
P value for H_0: $B_0 = 0$	0.7544	0.5986
Slope coefficient B_1	411	435
P value for H_0: $B_1 = 0$	0.0001	0.0054

Note: x = subplot area sampled (ha); y = number of subplot trees tallied.

variance of y_i is proportional to x_i^2. However, the mean of ratios often has a severe bias and Ek (1971) recommends against using it in forest surveys. In addition, difficulties occur when no condition classes of interest occur on a given plot. For example, consider estimating trees per hectare for loblolly pine when some plots are located entirely in non-loblolly pine types. In this case the observation y_i/x_i would be 0/0, an undefined quantity. A further complication resulting from this scenario relates to sample size (n). It is not clear if n should be reduced for such plots, thus creating a problem that has a direct effect on the variance calculations.

For all practical purposes such difficulties are eliminated with the ROM estimator. Equation 1 is still defined when any pair y_i and x_i are both equal to zero. In addition, the estimated variance (eqs. 2 and 3) is virtually unaltered by these observations when n is greater than 30 (which is one of the stated assumptions). This is easy to show by using eq. 3 and noting that $n(n - 1)$ is approximately equal to n^2 when n is greater than 30. Thus, substituting n^2 for $n(n - 1)$ in eq. 3 yields

$$\hat{V}(\hat{R}) = \frac{1}{n^2\left(\sum_{i=1}^{n} x_i/n\right)^2}\left(\sum_{i=1}^{n} y_i^2 + \hat{R}^2\sum_{i=1}^{n} x_i^2 - 2\hat{R}\sum_{i=1}^{n} y_ix_i\right)$$

$$= \frac{1}{\left(\sum_{i=1}^{n} x_i\right)^2}\left(\sum_{i=1}^{n} y_i^2 + \hat{R}^2\sum_{i=1}^{n} x_i^2 - 2\hat{R}\sum_{i=1}^{n} y_ix_i\right)$$

Hence, $\hat{V}(\hat{R})$ as written above is defined for x_i and y_i equal to zero, and n does not appear in the equation except as an

Fig. 2. Scatterplot of the number of trees (y) against plot size (x) for the FHM mapped-plot data from North Carolina.

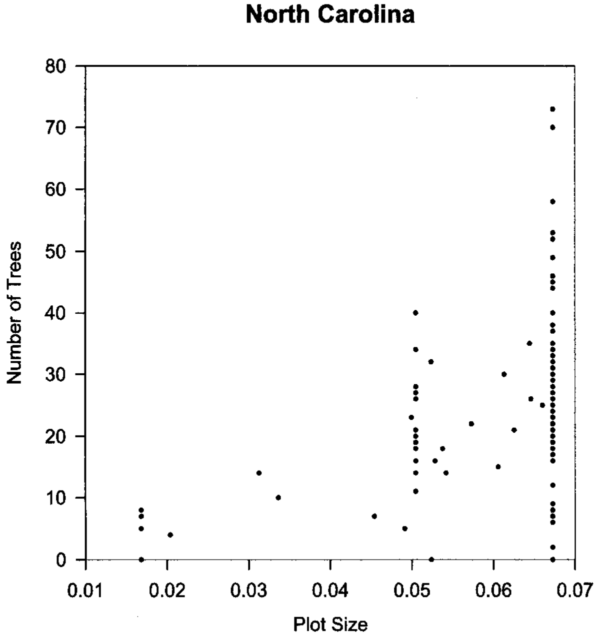
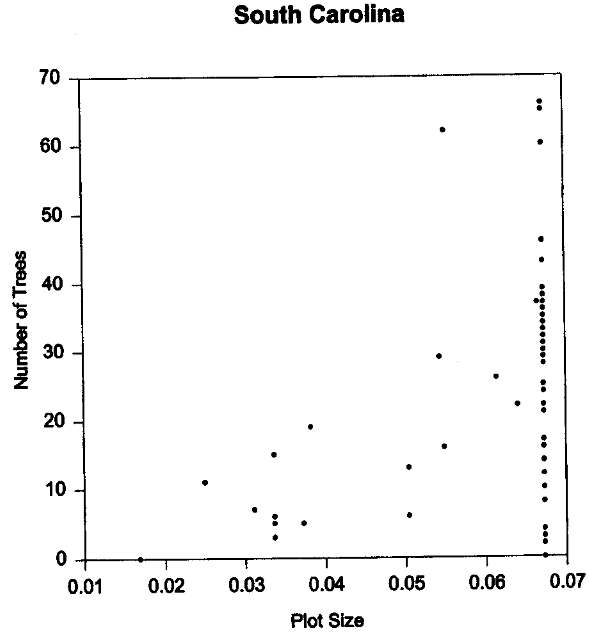


Fig. 3. Scatterplot of the number of trees (y) against plot size (x) for the FHM mapped-plot data from South Carolina.



index of summation, which solves both problems described above.

It is interesting to note that when plot size is constant over all plots, the ROM reduces to the MOR. If $x_i = x$, eq. 1 becomes

$$\hat{R}_{\text{ROM}} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x} = \frac{\sum_{i=1}^{n} y_i}{xn} = \frac{\bar{y}}{x}$$

and, since $S_{yx} = 0$ and $S_x^2 = 0$, eq. 2 becomes

$$\hat{V}(\hat{R}_{\text{ROM}}) = \frac{1}{n\bar{x}^2}(S_y^2 + \hat{R}^2 S_x^2 - 2\hat{R}S_{yx}) = \frac{1}{\bar{x}^2}\frac{S_y^2}{n}$$

When the constant $x_i = x$ is substituted into the MOR estimator, we obtain

$$\hat{R}_{\text{MOR}} = \frac{\sum_{i=1}^{n} \frac{y_i}{x}}{n} = \frac{1}{x}\bar{y} = \frac{\bar{y}}{x}$$

and

$$\hat{V}(\hat{R}_{\text{MOR}}) = \frac{1}{x^2}V(\bar{y}) = \frac{1}{\bar{x}^2}\frac{S_y^2}{n}$$

which are both equivalent to the ROM.

Weighted mean of ratios (WMOR)

When the MOR estimator is weighted by plot size, it reduces to the ROM whether plot size is constant or not. This is easy to see since

$$\hat{R}_{\text{WROM}} = \sum_{i=1}^{n} \frac{x_i}{\sum_{i=1}^{n} x_i}\frac{y_i}{x_i} = \sum_{i=1}^{n} \frac{y_i}{\sum_{i=1}^{n} x_i} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} = \hat{R}_{\text{ROM}}$$

However, treating the weight as a constant when deriving the variance is incorrect since it does not include this additional source of variation. The ROM properly handles this in its variance estimator.

The regression estimator

The regression estimator is appropriate when the relationship between y_i and x_i is linear (not necessarily through the origin) and the variance of y_i is homogeneous (a constant) across all x_i. Inspection of Table 2 indicates that this assumption does not hold up for the mapped design.

Nonclassical variance approximations

As opposed to the classical approaches described above, nonclassical variance approximations such as bootstrapping might be applied as an alternative method to estimate the variance of any of the estimators described above, including ROM. However, the benefits of choosing such a method are unclear and somewhat philosophical. In essence, the mapped-plot design is a two-stage sampling design where the plots are the primary sampling units and the subplots and microplots are the secondary sampling units. Within each primary sampling unit, the unequal-sized secondary sampling units consist of elements defined as either trees when tree-level estimates are derived or hectares when per-hectare estimates are derived. Since ROM estimators are appropriate for cluster sampling, we preferred a simpler traditional approach over nonclassical methodology.

Conclusions

High percentages of plots with multiple conditions show that the forest resource is highly fragmented across the U.S., confirming that the extra complexity incurred with the mapped-plot design is not spent on isolated anomalies. Of the numerous estimators that could be applied to this design, ROM appears qualified for the best linear unbiased estimator for inventory attributes involving relationships between trees and plot size. ROM also has several practical advantages over alternative methods. Occasionally a situation might be encountered where ROM is inferior to another estimator, but with large-scale inventories it is more practical (in terms of consistency, error reduction, and ease of processing) to utilize the single estimator that is usually the best, and forego the minor improvement that would be gained in a few isolated cases.

References

Brewer, K.W.R. 1963. Ratio estimation in finite populations: Some results deducible from the assumption of an underlying stochastic process. Aust. J. Stat. **5**: 93–105.

Cochran, W.G. 1977. Sampling techniques. 3rd ed. John Wiley & Sons, Inc., New York.

Ek, A.R. 1971. A comparison of some estimators in forest sampling. For. Sci. **17**: 2–13.

Fieller, E.C. 1932. The distribution of the index in a normal bivariate population. Biometrika, **24**: 428–440.

Hahn, J.T., MacLean, C.D., Arner, S.L., and Bechtold, W.A. 1995. Procedures to handle FIA cluster plots that straddle two or more conditions. For. Sci. Monogr. **31**: 12–25.

James, A.T., Wilkinson, G.N., and Venables, W.N. 1974. Interval estimates for a ratio of means. Sankhya Ser. A, **36**(Part 2): 177–183.

Kish, L. 1965. Survey sampling. John Wiley & Sons, Inc., New York.

Milne, A. 1959. The centric systematic area-sample treated as a random sample. Biometrics, **15**: 270–297.

Paulson, E. 1942. A note on the estimation of some mean values for a bivariate distribution. Ann. Math. Stat. **13**: 440–444.

Royall, R.M. 1970. On finite population sampling theory under certain linear regression models. Biometrika, **57**: 377–387.

Scott, C.T. 1993. Optimal design of a plot cluster for monitoring. *In* The optimal design of forest experiments and forest surveys, September 10–14, 1993, School of Mathematics, Statistics and Computing, University of Greenwich, London. pp. 233–242.

Scott, C.T., and Bechtold, W.A. 1995. Techniques and computations for mapping plot clusters that straddle stand boundaries. For. Sci. Monogr. **31**: 46–61.

Scott, C.T., Ek, A.R., and Zeisler, T.R. 1983. Optimal spacing of plots comprising clusters in extensive forest inventories. *In* Renewable inventories for monitoring changes and trends. Society of American Foresters, Bethesda, Md. SAF 83-14: 707–710.